

Accordingly, the variance-covariance matrix of the phenotype can be expressed as:

$$\text{Var}(\mathbf{y}) = \sigma_g^2 \mathbf{G} + \sigma_e^2 \mathbf{I}$$

Under this model, heritability is estimated as:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$$

This framework provides the theoretical foundation for GREML (genomic-relatedness-based restricted maximum likelihood), enabling the estimation of additive genetic variance explained by genome-wide markers through statistical inference (Da et al., 2014; Yang et al., 2016; Zhou et al., 2020). Furthermore, to better capture complex genetic architectures, extensions of the LMM have been proposed, such as models incorporating multiple random effects or covariance structures among random effects (Zhou et al., 2019; 2020).

#### 4.2 REML estimation and the maximum likelihood framework

In terms of parameter estimation, GREML typically relies on restricted maximum likelihood (REML). Unlike conventional maximum likelihood (ML), REML eliminates fixed effects by integrating them out of the likelihood function, thereby optimizing variance parameters based on residuals. This approach effectively avoids bias in variance component estimation caused by fixed-effect estimation, and is particularly advantageous in complex models and finite-sample settings (Dao et al., 2021; Meyer, 2023).

In practical implementation, REML is carried out via numerical optimization of the log-likelihood function. The GCTA software employs the AI-REML (Average Information REML) algorithm, which iteratively updates parameters using the average information matrix and achieves efficient estimation of variance components (Yang et al., 2016; Strandén et al., 2024). BOLT-REML introduces stochastic projection and approximation techniques to substantially reduce computational complexity in large-scale datasets, making it suitable for cohorts with sample sizes on the order of hundreds of thousands to millions (Border and Becker, 2019). The GEMMA software also implements the REML framework and extends it to multivariate and Bayesian analyses, demonstrating robust convergence properties in small to medium-sized datasets (Meyer, 2023).

Recent methodological advances, including principal component-based reparameterization and stochastic optimization algorithms, have further improved the scalability and adaptability of REML estimation for large and complex datasets (Strandén et al., 2024).

#### 4.3 Validation using simulated and empirical data

The validity of the GREML method is typically assessed through a combination of simulation studies and empirical data analyses. Simulation studies have shown that, under correct model specification and sufficiently large sample sizes, GREML can provide unbiased estimates of heritability (Da et al., 2014; Cesarani et al., 2018; Zhou et al., 2020). However, in small sample settings (e.g., hundreds of individuals), the limited information content of the GRM leads to large estimation variance, and the estimates become sensitive to assumptions regarding population structure and phenotypic distribution, potentially introducing bias (Cesarani et al., 2018; Meyer, 2023).

In contrast, in large cohorts (tens of thousands to millions of individuals), GREML is capable of more accurately capturing the genetic variation explained by genome-wide markers. Approximate methods such as BOLT-REML have been shown, in human population studies (e.g., UK Biobank), to produce heritability estimates close to true values while effectively controlling for population structure and batch effects (Nolte et al., 2017; Ni et al., 2018). In crop populations, such as maize and wheat with genome-wide data, GREML applications have revealed the heritable architecture of complex quantitative traits and provided theoretical guidance for subsequent GWAS and genomic selection. Further methodological extensions, such as CORE GREML, allow for covariance among random effects and have demonstrated improved performance over standard GREML in the presence of complex genetic architectures (Zhou et al., 2019; 2020).